To: Unicode Technical Committee
From: Azzeddine Lazrek, Professor - Cadi Ayyad University Morocco,
　　　　Consultant - King Fahd Glorious Quran Printing Complex Saudi Arabia
Contact: lazrek@uca.ac.ma, aab@qurancomplex.gov.sa,
　　　　http://alhabl.uca.ma/, http://www.qurancomplex.org/
Subject: Comments about "*Arabic additions for Quranic orthographies*" proposition
Date: October 30, 2019

We would like to share here some remarks and comments about "*Arabic additions for Quranic orthographies*", the very important proposition from Roozbeh Pournader and Deborah Anderson (L2/19-306). We hope this will contribute to properly preserve the Arabic Unicode model.

1. ARABIC SMALL HIGH WORD SAH
Comment: Correct proposal.

2. ARABIC SYMBOL WASLA ABOVE
Comment: For this character, it will be better to consider it as a combining character, not only for pedagogical use, but above all, for a correct coding following the Unicode model (see L2/18-028).
Effectively, we realize that this will generate two different ways to represent the existing Unicode character U+0671 ARABIC LETTER ALEF WASLA. However, that poses fewer problems, compared to the important advantage that will be ensured as we will exemplify (see characters from 8 to 26 below). Mostly, there is an equivalence relationship between them.
We think that coding the existing Unicode character U+0671 was not a good decision in past. Moreover, it is not the first time to have two different ways to represent a character at Unicode, even if we must take care to avoid multi coding.

3. ARABIC SMALL HIGH FARSI YEH
Comment: For this character, we propose to change its name to be: "ARABIC SMALL HIGH DOTLESS YEH", as for: U+066E ARABIC LETTER DOTLESS BEH, U+066F ARABIC LETTER DOTLESS QAF, U+06A1 ARABIC LETTER DOTLESS FEH (see *Michael S. Kaplan* http://archives.miloush.net/michkap/archive/2006/02/14/531572.html).

4. ARABIC SMALL FARSI YEH
Comment: For this character, we propose to change its name to be: "ARABIC SMALL DOTLESS YEH", as preceding one.

5. ARABIC SUPERSCRIPT ALEF MOKHASSAS
6. ARABIC BASELINE ROUND DOT
7. ARABIC RAISED ROUND DOT
Comment: Correct proposals.

8. ARABIC LETTER ALEF WITH ATTACHED FATHA
9. ARABIC LETTER ALEF WITH RIGHT MIDDLE STROKE
10. ARABIC LETTER ALEF WITH ATTACHED KASRA
11. ARABIC LETTER ALEF WITH ATTACHED TOP RIGHT FATHA
12. ARABIC LETTER ALEF WITH LEFT MIDDLE STROKE
13. ARABIC LETTER ALEF WITH ATTACHED BOTTOM RIGHT KASRA
14. ARABIC LETTER ALEF WITH ATTACHED ROUND DOT ABOVE

15. ARABIC LETTER ALEF WITH ATTACHED LEFT ROUND DOT
16. ARABIC LETTER ALEF WITH ATTACHED RIGHT ROUND DOT
17. ARABIC LETTER ALEF WITH ATTACHED ROUND DOT BELOW
18. ARABIC LETTER ALEF WITH DOT ABOVE
19. ARABIC LETTER ALEF WITH ATTACHED TOP RIGHT FATHA AND DOT ABOVE
20. ARABIC LETTER ALEF WITH RIGHT MIDDLE STROKE AND DOT ABOVE
21. ARABIC LETTER ALEF WITH ATTACHED BOTTOM RIGHT KASRA AND DOT ABOVE
22. ARABIC LETTER ALEF WITH ATTACHED TOP RIGHT FATHA AND LEFT RING
23. ARABIC LETTER ALEF WITH RIGHT MIDDLE STROKE AND LEFT RING
24. ARABIC LETTER ALEF WITH ATTACHED BOTTOM RIGHT KASRA AND LEFT RING
25. ARABIC LETTER ALEF WITH ATTACHED RIGHT HAMZA
26. ARABIC LETTER ALEF WITH ATTACHED LEFT HAMZA

Comment: For the signs, present in these characters, numbered from 8 to 26, that come with the letter Alef - according to the various Quranic readings and as it was noted in some countries at some times -, we agree to code them, but we still object only to the method of their coding. These signs must be considered as combining characters without the letter Alef, although they come only with it, or alone for pedagogical purpose; as it was already explained the scientific reason for this objection about the coding method, not the coding objection itself, in L2/18-028.

In the same way, we must avoid the same mistake already present for the only existing Unicode character U+0671 ARABIC LETTER ALEF WASLA. For our opinion, Wasla sign must be a combined character without the letter Alef, even if it is present only with it (see character 2 above).

This method of coding is not at all coherent with the process of coding in general, and with the principles of Unicode coding in particular.

This error will now be repeated with the letters numbered from 8 to 26, if the signs are coded as letters with Alef, and are not independent from it.

As it is known, text coding is very important for exploiting it in processing, sharing, searching, etc.

For more efficient codification, we must see codified characters as they are in reality. In fact, in the Arabic script, generally, we must consider two types of information that a letter may contain: the letter itself, and the additional information to that letter. For example, the diacritical marks on a letter will not change it to another letter, i.e. "بَ" and "بُ" designate the same letter Beh, except the diacritic sign that changes. Similarly, returning to the mentioned characters, we find that all of it designate the same letter: Alef as Hamzat Al Wasl (همزة الوصل), which is used in the Arabic script at the beginning of words starting with a letter having the diacritic Sukun "ْ". What differ in these different cases is its original diacritic (marked by the dot at the top, bottom or middle) and/or the diacritic of the letter which precedes it (marked by the stroke at the top, bottom or middle). The difference between the two proposed codification methods (our proposition against L2/19-306 once) manifests more in text processing. For example, if we want to remove all the diacritics and leave only the letters, at L2/19-306′ proposition we will have a problem since the diacritical mark is merged with its letter; however, at our proposition we will have no problem since the diacritical mark is not merged with its letter but is a combining character without letter.

Furthermore, when searching for texts containing the word "اعبدوا", for example, search engines must find all texts and Quranic-texts in which this word is mentioned, regardless of

the presence or absence of idiomatic marks for special Quranic readings, at all printed Mushafs. But by doing a search (manually) of the word "اعبدوا" in the Quran (Warsh version), we find that there are 6 cases where the Hamzat Al Wasl is preceded by Damma, 5 cases where it is preceded by Fatha and 10 cases where it is preceded by a Kasra; while the search engine must recognize all the 21 cases at the same time.

To my humble knowledge, we will not be able to get this important result if the method of coding proposed right in L2/19-306 is adopted. Thus we will not get the desired benefit from the general coding process, and from the Unicode coding in particular.

Fortunately, in Unicode, all Arabic diacritical marks are coded as combining characters without Arabic letters; which will allow having them alone for a pedagogical purpose at any word processor, simply by composing them with a space character.

Note that one of the principles of Unicode is to not delete any character after adoption, even if it turns out after making a mistake, as happened in the coding of Arabic letters with contextual forms, its four different glyphs according to their location in the word (as the 4 glyphs: "ج" "جـ" "ـجـ" "ـج" for Arabic letter Jeem) (Arabic Presentation Forms-B Range: FE70–FEFF).

For the chosen names, it is used **FATHA** for Fatha, **KASRA** for Kasra but **MIDDLE STROKE** for Damma; which is not homogenized. We propose to always use **STROKE** for all the 3 cases.

In another hand, we should not talk about attachment of sign **STROKE** with Alef. Indeed, it is not observed in all printed Mushafs, as in the following samples already presented in L2/17-252. Moreover, it is a font purpose. The term **ATTACHED** should not be used in the name. If we use the term **ATTACHED** in the name to code those where attachment is observed, we will have to code others for those whose attachment is not observed at all, for same semantic purpose.
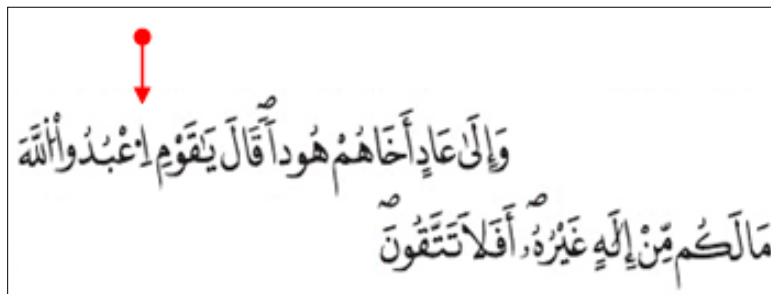

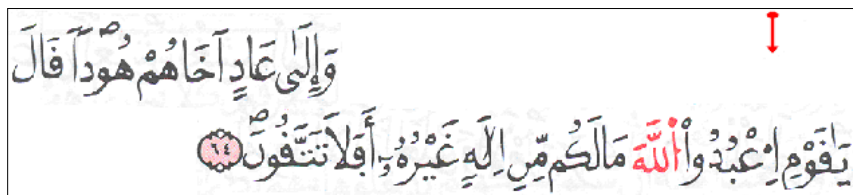[Mushaf Warsh, Dar AlMushaf, Beirut, pp. 520]
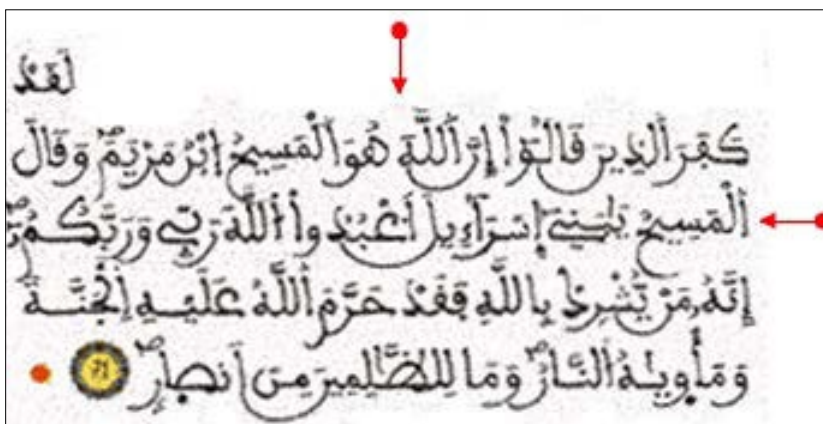
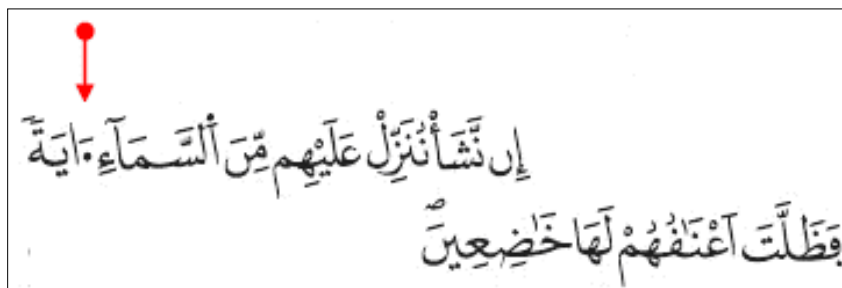[Mushaf Hassani, Fdalat Printer, Mohamadia, 2009, pp. 127]
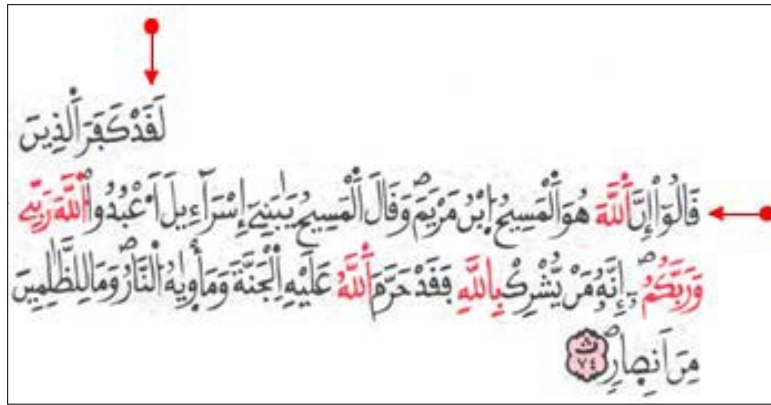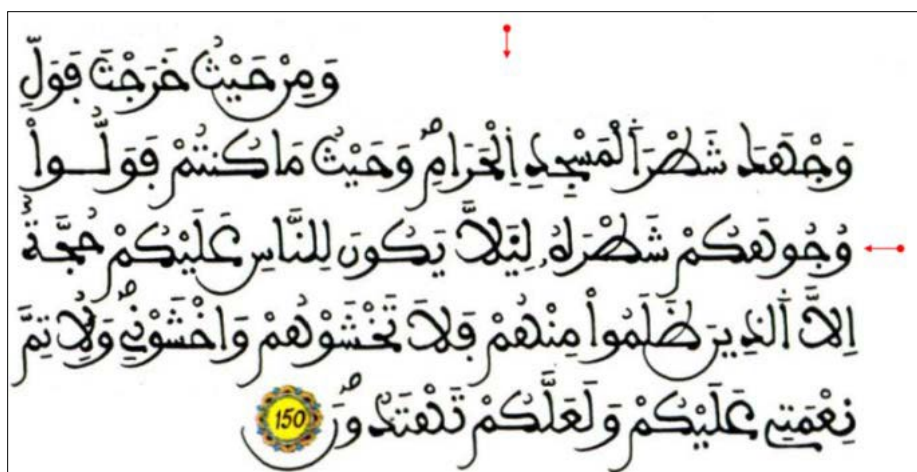
[Mushaf Warsh, Dar AlMushaf, Beirut, pp. 88]

[Mushaf AlMadinah Anabawiyat, Riwayah Qaloon, King Fahd Glorious Quran Printing Complex, 2005, pp. 138]

[Mushaf Warsh, Dar AlMushaf, Beirut, pp. 122]

[Mushaf AlMadinah Anabawiyat, Riwayah Warsh, King Fahd Glorious Quran Printing Complex, 2005, pp. 328]

[Mushaf Hassani, Fdalat Printer, Mohamadia, 2009, pp. 24]

In Unicode, the combining diacritical marks, as accents for Latin script and Fatha, Kasra, etc. for Arabic script, are combining characters.

The Unicode Arabic model follows until now works very well. It allows a base character, for all letters such as Alef, and a combining character, for all diacritical marks such as Fatha. The proposed precomposed characters, numbered from 8 to 26, will not follow the Unicode model. As signs present on those characters are also diacritical marks, they must be coded as combining characters.

The only exception was the existing Unicode character U+0671 ARABIC LETTER ALEF WASLA. We can fix this problem only by consider the proposed characters, number 2: **ARABIC SYMBOL WASLA ABOVE**, as a combining character.

Our proposition of coding method for these characters, numbered from 8 to 26, is already presented in detail in L2/17-252.
In our way, the display process will not be modified, the font enlargement will be easy and mainly the search engine will be efficient.

27. ARABIC SUKUN BELOW
Comment: For this character, we propose to change its name to be: "**ARABIC SMALL CIRCLE BELOW**" as there is no relationship with the well known Sukun sign.

28. ARABIC LARGE CIRCLE BELOW
29. ARABIC ROUND DOT INSIDE LARGE CIRCLE BELOW
30. ARABIC DOUBLED MADDA
31. ARABIC HALF MADDA OVER MADDA
32. ARABIC SMALL HIGH ZAH
33. ARABIC LETTER THIN YEH
34. ARABIC TATWEEL WITH TWO DOTS BELOW
35. ARABIC TATWEEL WITH OVERSTRUCK WAW
36. ARABIC TATWEEL WITH OVERSTRUCK HAMZA
37. ARABIC SMALL HIGH YEH BARREE WITH TWO DOTS BELOW
38. ARABIC LARGE ROUND DOT ABOVE
39. ARABIC LARGE ROUND DOT BELOW
Comment: Correct proposals.